REMARKS

This is responsive to an Office Action mailed on August 4, 2009. The Office Action rejected claims 1-5, 7-12, 14-27 and 29-31. Applicant has amended claims 1, 14 and 25. The application currently includes claims 1-5, 7-12, 14-27 and 29-31.

The Office Action rejected claims 1-5 and 7-12 under 35 USC §112, second paragraph. Applicant respectfully submits that claim 1 has been amended to overcome the 35 USC § 112, second paragraph rejection and respectfully requests that the rejections of claims 1-5 and 7-12 be withdrawn.

The Office Action reports that claims 1-4, 6-7, 14-21, 23, 25-26, and 28 were rejected under 35 U.S.C. §103 as being unpatentable over U.S. Patent No. 5,806,021 to Chen et al. (hereinafter "Chen") in view of U.S. Patent No. 6,968,308 to Brockett et al. (hereinafter "Brockett") It is respectfully submitted that the cited references even when combined do not teach or suggest all of the features of claim 1.

Applicants respectfully submit that independent claim 1, as well as their dependent claims are allowable over the cited prior art. Claim 1 includes utilizing FMM and BMM to tokenize the sentence. The tokenized sentence of characters is classified into known characters and at least one overlapping ambiguity string, wherein the overlapping ambiguity string comprises at least three Chinese characters having at least two possible segmentations, wherein each possible segmentation comprises a right portion and a left portion and wherein the left portion and the right portion remain in the tokenized corpus and the at least one overlapping ambiguity string is removed from the tokenized corpus.

The Chen reference in combination with the Brockett reference does not teach, suggest or render obvious the claimed method. The Chen reference explicitly discloses utilizing **either** a forward/backward matching segmenter **or** a statistical stack search segmenter. Utilizing the forward/backward matching method, word boundaries are located by forming the longest words which exist in the lexicon (i.e., the longest allowable word), in both forward and backward manners. A statistical language model is utilized then to make a decision when there is any output discrepancies between the forward and backward matching. Col. 2, lines 39-44 and Col. 3,

lines 18-33. Applicants submit that probabilities are based upon the statistical probability disclosed in the Chen patent is that of a word being utilized in the Chinese language, and not the probability based on context as claimed. There is no disclosure in the Chen patent of utilizing the context of the words adjacent the output discrepancies to determine a probability related to context of a possible segmentation as claimed.

The Chen patent, having established rules for forward/backward segmenting, would not need to utilize context of the phrases to determine a probability because the probabilities have already been established independent of context. There is no disclosure of using at least one context feature comprising a Chinese character to determine the probability of whether one segmentation is more likely compared to an alternative segmentation. Therefore, Applicants respectfully submit that a set of rules is established with respect to utilizing forward/backward searching in the Chen patent and that a context feature is not considered in determining the probability.

Also the Chen patent does not disclose retaining the left portion and the right portion of the overlapping ambiguity string while the at least one overlapping ambiguity string is removed from the tokenized corpus as claimed. Rather, the Chen patent discloses that the overlapping ambiguity string is retained and a statistical language model is utilized to determine a selected output.

Applicants submit that the Brockett patent does not cure the deficiencies of the Chen patent. Rather, Brockett discloses the use of tries to determine all of the possibilities of these characters in forming words. Each of these words is then given a value and if the combined characters are not part of the sentence then the combination is removed from the analysis. The combination of characters with the highest word probability is then assigned to the sentence. There simply is no disclosure in the Brockett patent of using context including at least one Chinese character to determine the probability of a combination of character segments would be most probable.

Applicant respectfully submits that the Brockett patent does not disclose utilizing forward and backward maximum matching searches or techniques to determine an overlapping ambiguity string. Rather, the Brockett patent discloses only a forward search engine where the ambiguity is determined only on a partial overlapping ambiguity string. As stated in the Office Action, a

sentence represents characters ABCDE and there are two possible segmentations, ABCD and BCDE. This is not one ambiguity string as claimed in the claimed invention. Rather, a portion of one string may also be overlapped with a portion of another string and whereby the Brockett patent discloses utilizing a sentactic parser to determine which string should be utilized.

This is not the same as what is claimed in the present invention. The present invention provides at least three letters which may have two separate interpretations. For instance, the string ABC may be A/BC or AB/C. As such, Applicant respectfully submits that the Brockett patent along with the Chen patent does not render obvious the claimed invention.

For instance, in column 11, lines 5-19, a portion of which was alleged to make independent claim 1 obvious, it is stated that the syntactic parser 316 produces a single parse at its output. A single parse does not teach, suggest or render obvious an FMM and a BMM as claimed. Further, one would not need to utilize the syntactic parser as disclosed in the Brockett patent with the claimed invention because they are utilized for different purposes. Namely, the Brockett patent discloses utilizing the syntactic parser when you have a portion of a phrase that overlaps with a portion of another phrase where the two phrases are different have different elements. In contrast, the claimed invention provides a method of determining the most likely output of a common phrase based upon probability information related to the context for each possible segmentation.

As such, Applicant respectfully submits that claim 1 and its dependent claims 2-5 and 7-12 are allowable over the prior art. Reconsideration and allowance of claims 1-5 and 7-12 are respectfully requested.

Claims 14 and 25 differs from claim 1 in that claims 14 and 25 claim determining constituent lexical words in the overlapping ambiguity string instead of left and right portions. Claim 25 also claims utilizing an N-gram model to obtain probability information along with contextual information including adjacent words.

For the reasons stated with respect to the allowability of claim 1, claims 14 and 25 are also allowable over the prior art. Reconsideration and allowance of claims 14 and 25 and dependent claims 15-24, 26, 27 and 29-31 are respectfully requested.

The foregoing remarks are intended to assist the Office in examining the application and in the course of explanation may employ shortened or more specific or variant descriptions of some of the claim language. Such descriptions are not intended to limit the scope of the claims; the actual claim language should be considered in each case. Furthermore, the remarks are not to be considered exhaustive of the facets of the invention which are rendered patentable, being only examples of certain advantageous features and differences, which applicant's attorney chooses to mention at this time. For the foregoing reasons, applicant reserves the right to submit additional evidence showing the distinction between applicant's invention to be unobvious in view of the prior art.

Furthermore, in commenting on the references and in order to facilitate a better understanding of the differences that are expressed in the claims, certain details of distinction between the same and the present invention have been mentioned, even though such differences do not appear in all of the claims. It is not intended by mentioning any such unclaimed distinctions to create any implied limitations in the claims.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to Deposit Account No. 23-1123.

Respectfully submitted,

MICROSOFT CORPORATION

By: /Peter J. Ims/
Peter J. Ims, Reg. No. 48,774
Westman, Champlin & Kelly, P.A.
900 Second Avenue South, Suite 1400
Minneapolis, MN 55402-3244

One Microsoft Way
Redmond, Washington 98052-6399
Phone: (425) 707-9382

PJI/mek